Reproduce Stitch3D

MA Xiaoheng
Department of Mathematics, HKUST
STUDENT ID: 21024750

December 15, 2024

1 Task Description

1.1 Background

Spatial transcriptomics (ST) technologies have revolutionized our understanding of tissue architecture by enabling high-throughput gene expression profiling while preserving spatial information in intact tissues. Recent technological developments have enabled the generation of ST datasets comprising multiple parallel 2D slices (x-y plane) along the z-axis within a tissue sample. These multi-slice datasets present unprecedented opportunities to construct comprehensive 3D representations of biological systems which can help us to interpret biological processes that naturally occur in three-dimensional (3D) space. However, current analytical approaches predominantly focus on 2D analysis within individual tissue sections, often leading to inconsistent interpretations and failing to capture the full complexity of spatial biological processes.

1.2 Problem Statement

The fundamental challenge is to reconstruct 3D cellular structures by integrating multiple 2D ST slices. This involves two primary tasks:

1.2.1 3D Spatial Domain Identification

The first task is to identify biologically interpretable 3D spatial regions where spots have similar gene expressions. This is crucial as it reveals comprehensive tissue structures in 3D space, enabling downstream analyses such as detection of region-related genes with 3D spatial patterns. Furthermore, it helps understand how biological processes influence the organization of different cell types. Through this spatial domain identification, we can better comprehend the complex three-dimensional architecture of biological tissues and their functional organization.

1.2.2 3D Cell-type Deconvolution

The second task is to infer 3D fine-grained cell-type distributions by integrating multiple ST slices with single-cell RNA-sequencing (scRNA-seq) atlases. This step is necessary because current ST technologies, while capable of detecting transcriptome-wide gene expressions within spatial spots, often capture multiple cells within each spot. By decomposing these cell-type

mixtures in spatial spots, we can achieve higher resolution for 3D reconstruction. This enhanced resolution allows for deeper insight into the biological functions of specific cell-type-enriched areas, providing a more detailed understanding of tissue organization and function.

2 STitch3D Algorithm

2.1 Overview

STitch3D is a deep learning-based method that reconstructs 3D tissue structures from multiple 2D spatial transcriptomics (ST) slices. The algorithm integrates multiple ST slices and a matched single-cell RNA sequencing (scRNA-seq) reference through a graph attention network architecture. The preprocessing steps first align these ST slices to obtain aligned 2D coordinates, which are then used to construct 3D coordinates and subsequently build a 3D adjacency graph. This differs from methods like STAligner, which constructs 2D adjacency graphs within individual slices. Since STAligner does not perform slice alignment beforehand, constructing 3D adjacency graphs based on original coordinates would be unreasonable. Following preprocessing, the data is integrated using the STitch3D model to obtain low-dimensional representations of each spot in a shared latent space, which are used for spatial domain identification and cell type deconvolution. This latent space is designed to preserve meaningful biological distinctions while mitigating batch effects. The final outputs include low-dimensional representations and cell type proportions for each spot, enabling comprehensive 3D tissue structure reconstruction. Through this systematic approach, STitch3D provides a robust framework for analyzing and understanding complex spatial transcriptomics data in three dimensions.

2.2 Input Processing

1. **Input Data:**

   - Assume there are S spatial transcriptomics slices, indexed by $s = 1, 2, \ldots, S$. The number of spots in the s-th slice is N_s, and the gene expression matrix is $Y^s = [Y^s_{n,g}] \in \mathbb{R}^{N_s \times G}$, where $n = 1, 2, \ldots, N_s$ is the spot index, and $g = 1, 2, \ldots, G$ is the gene index.
   - In the corresponding scRNA-seq reference, the gene expression matrix for different cell types is $V = [V_{c,g}] \in \mathbb{R}^{C \times G}$, where $c = 1, 2, \ldots, C$ is the cell type index. Each row in the matrix represents the mean expression profile of a specific cell type, with the constraint $\sum_g V_{c,g} = 1$.

2. **Preprocessing Steps:**

   - Align multiple slices using ICP or PASTE algorithms to establish 3D spatial coordinates, where z-coordinates represent physical distances between slices
   - Construct global 3D neighborhood graph where spots are considered adjacent if their distance is less than 1.1 times the minimum intra-slice spot distance
   - Concatenate gene expression matrices $Y = [Y_{i,g}] \in \mathbb{R}^{N \times G}$ from S slices sequentially and reindex all spots from 1 to N, where $N = N_1 + N_2 + \ldots + N_s$ represents the total number of spots

2.3 Core Model Components

2.3.1 Latent Space Encoding

The gene expression data is first encoded into a shared latent space:

$$X_{i,g} = \log\left(\frac{Y_{i,g}}{\sum_{g=1}^{G} Y_{i,g}} \times 10^4 + 1\right) \tag{1}$$

$$Z = f_Z(X, A) \tag{2}$$

where:

- X is the normalized and log-transformed gene expression
- A is the global adjacency
- $f_Z(\cdot)$ is a graph attention network
- $Z \in \mathbb{R}^{N \times p}$ represents the latent space embeddings, p represents the dimension of latent space.

2.3.2 Cell-type Proportion Estimation

Cell-type proportions are generated from latent representations using a neural network:

$$\beta_i = f_\beta(Z_i) \tag{3}$$

where:

- $\beta_i = [\beta_{i,1}, ..., \beta_{i,C}]^T$ represents the cell-type proportion vector for spot i. Here, $\beta_{i,c}$ denotes the proportion of cell type c in spot i. The proportions satisfy two constraints: $\sum_{c=1}^{C} \beta_{i,c} = 1$ (the sum of all cell-type proportions equals 1) and $\beta_{i,c} \geq 0$ (each proportion is non-negative).
- Z_i is the latent representation of spot i
- $f_\beta(\cdot)$ is a neural network.

This design allows the model to learn complex relationships between spatial patterns and cell-type distributions.

2.3.3 Batch Effect Modeling

Two effects are introduced to account for technical variations:

- $\alpha_i^s = f_\alpha(Z_i, s)$ represents slice/spot-specific effects.
- γ_g^s represents slice/gene-specific effects

2.4 Model Training

2.4.1 Count Data Modeling

The observed counts are modeled as:

$$Y_{i,g} \sim \text{Poisson}(l_i \lambda_{i,g}) \tag{4}$$

$$\lambda_{i,g} = \exp[\log(\sum_{c=1}^{C} \beta_{i,c} V_{c,g}) + \alpha_i^s + \gamma_g^s] \tag{5}$$

2.4.2 Loss Function

The main loss function is derived from the Maximum Likelihood Estimation (MLE) of a Poisson model. Given that gene expression counts are typically modeled as Poisson distributions:

$$Y_{i,g} \sim \text{Poisson}(l_i \lambda_{i,g}) \tag{6}$$

where $Y_{i,g}$ represents the observed gene expression count for spot i and gene g, l_i is the total transcript count in spot i, and $\lambda_{i,g}$ is the rate parameter modeled as:

$$\lambda_{i,g} = \exp\left[\log\left(\sum_{c=1}^{C} \beta_{i,c} V_{c,g}\right) + \alpha_i^s + \gamma_g^s\right] \tag{7}$$

The probability density function of Poisson distribution is:

$$P(Y_{i,g} | l_i \lambda_{i,g}) = \frac{(l_i \lambda_{i,g})^{Y_{i,g}} e^{-l_i \lambda_{i,g}}}{Y_{i,g}!} \tag{8}$$

Taking the negative log-likelihood and summing over all spots and genes:

$$-\log L = -\sum_{i=1}^{N} \sum_{g=1}^{G} [Y_{i,g} \log(l_i \lambda_{i,g}) - l_i \lambda_{i,g} - \log(Y_{i,g}!)] \tag{9}$$

Dropping the constant term and normalizing by N, we obtain the main loss function:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \sum_{g=1}^{G} [Y_{i,g} \log(l_i \lambda_{i,g}) - l_i \lambda_{i,g}] \tag{10}$$

To preserve biological variations across slices while handling batch effects, an autoencoder regularizer is introduced:

$$R_{AE} = \frac{1}{N} \sum_{i=1}^{N} \| f_X(Z_i, s) - X_i \|_2 \tag{11}$$

where $f_X(\cdot, s)$ is a slice-specific decoder network, Z_i is the latent representation, X_i is the normalized and log-transformed gene expression, and s is the slice label. This regularizer ensures that the latent space captures biological variations while the slice-specific decoder accounts for technical variations.

The final objective function combines both terms:

$$\text{Objective} = L + k_{AE} R_{AE} \tag{12}$$

where k_{AE} is the regularization coefficient set to 0.1.

3 Simulations

3.1 Methods

We implemented two versions of the model in this section:

- STitch3D: The complete model as described in the original paper, with batch effect correction terms α_i^s and γ_g^s
- STitch3D_no_batch_effect: A simplified version without batch effect correction using only:

$$\lambda_{i,g} = \exp[\log(\sum_{c=1}^{C} \beta_{i,c} V_{c,g})]$$

3.2 Experiments on Human Dorsolateral Prefrontal Cortex (DLPFC) Dataset

We first conducted some experiments on the human dorsolateral prefrontal cortex (DLPFC) dataset to evaluate the performance of STitch3D and compare it with STitch3D_no_batch_effect.

3.2.1 Slice Alignment and 3D Reconstruction

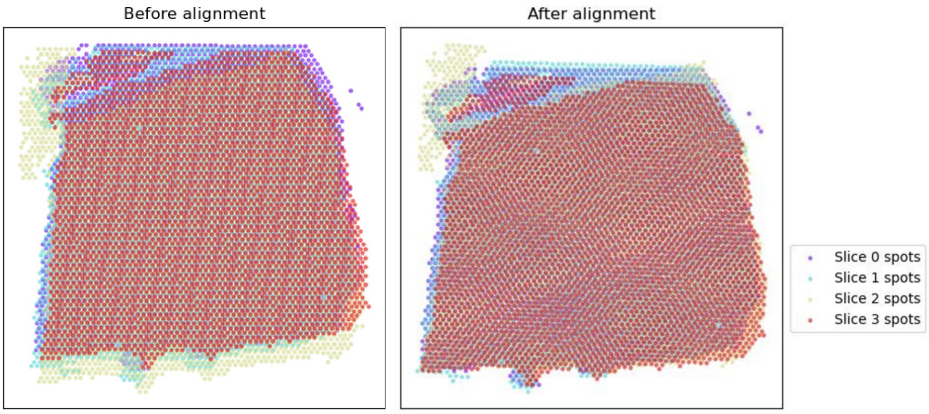

Figure 1: Comparison of spatial organization before and after slice alignment. (Left) Original slice positions. (Right) Aligned slice positions showing improved spatial correspondence.

First, we visualized the effectiveness of the slice alignment process by comparing the spatial organization of spots before and after alignment (Figure 1). The alignment procedure significantly improved the spatial correspondence between adjacent slices, providing a solid foundation for subsequent 3D reconstruction.

The reconstructed tissue slices (Figure 2) demonstrate marked differences between the two methods:

- STitch3D: Achieved coherent reconstruction with clear layer organization and smooth transitions between slices
- STitch3D_no_batch_effect: Showed discontinuities between slices and less consistent layer organization

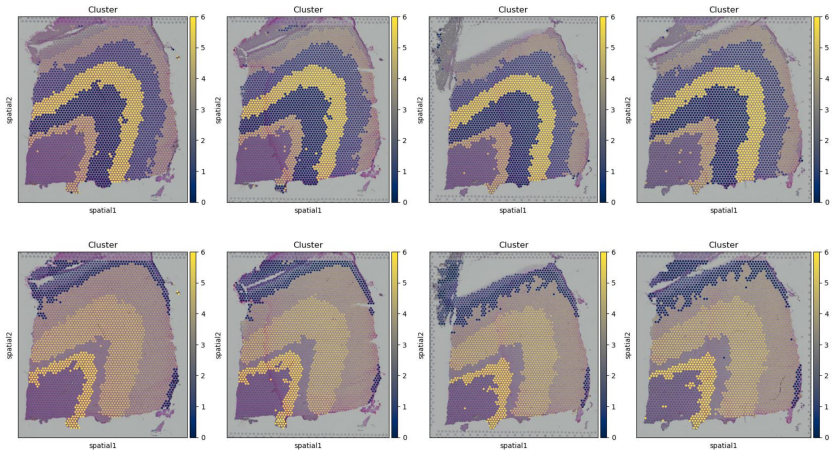

Figure 2: Reconstructed tissue slices using (Top Row) STitch3D and (Botton Row) STitch3D_no_batch_effect, demonstrating the impact of batch effect correction on reconstruction quality.

3.2.2 Cell-type Deconvolution Analysis

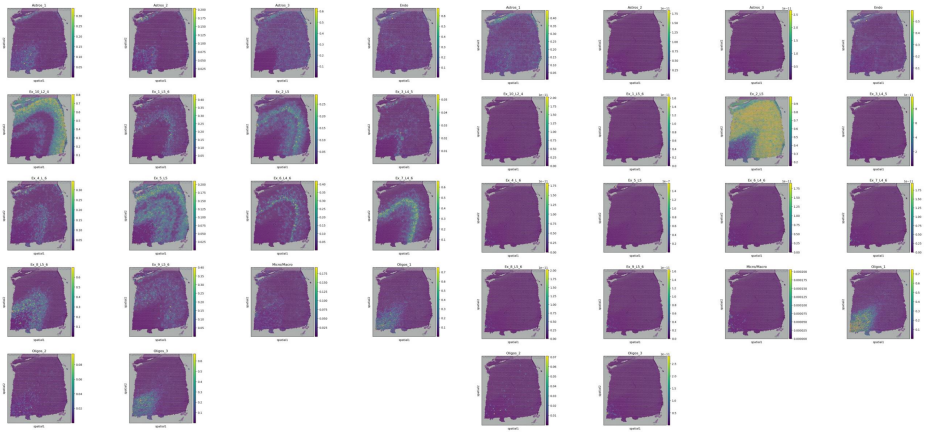

Figure 3: Cell-type deconvolution results of STitch3D(Left) and STitch3D_batch_effect(Right). Different colors represent distinct cell types, with intensity indicating proportion values.

Figure 3 presents a comprehensive comparison of cell-type deconvolution results between STitch3D and STitch3D_no_batch_effect methods. In these visualizations, different colors represent distinct cell types, while the color intensity indicates the proportion of each cell type at each spatial spot. The results clearly demonstrate that STitch3D successfully reconstructs the spatial distribution of cell types, with different cell populations showing distinct abundance patterns across various spots. In contrast, the method without batch effect correction shows inferior performance, likely due to inadequate batch effect removal. This is evidenced by the dominance of only a few cell types (such as Ex_2_L5) showing high proportions across numerous spots, while other cell types are rarely detected. This suggests poor reconstruction of

cellular spatial organization. The comparison highlights the superior capability of STitch3D in cell-type deconvolution analysis and the success of the batch effect correction method used in STitch3D.

3.2.3 Latent Space Analysis

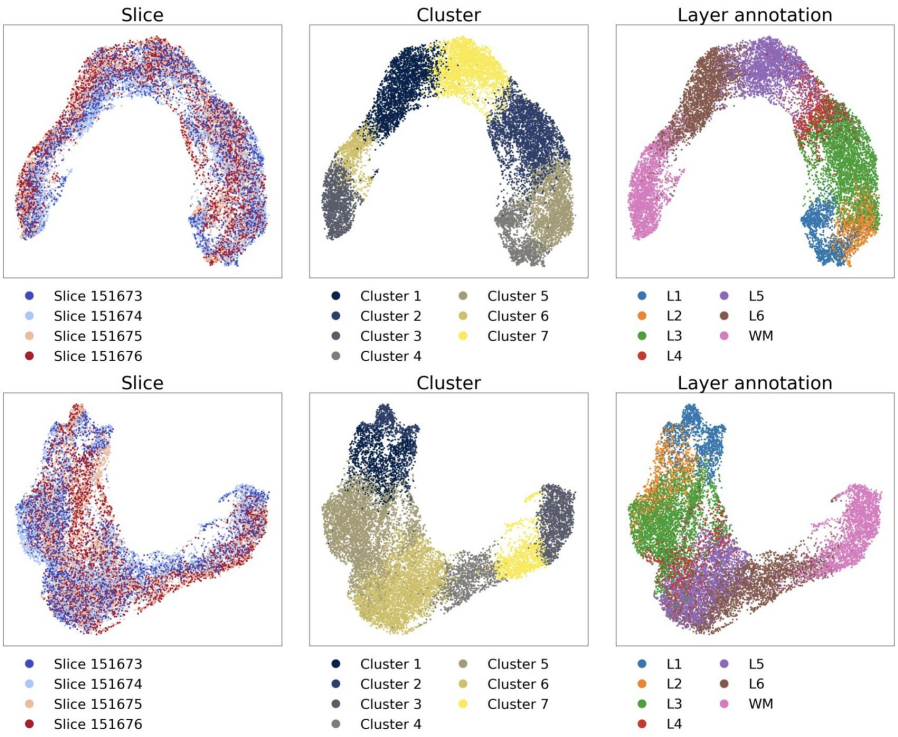

Figure 4: UMAP visualization comparing learned representations between models. Top row: STitch3D with batch effect correction. Bottom row: STitch3D_batch_effect. For each model, spots are colored by (left) slice indices, (middle) identified spatial domains, and (right) manual layer annotations.

We compared two versions of the model to demonstrate the importance of batch effect correction: The UMAP visualization (Figure 4) reveals striking differences between the two approaches. The complete STitch3D model demonstrates superior performance in several aspects. First, it successfully integrates data from different slices into a coherent latent space, as evidenced by the thorough mixing of spots from different slices (top left). Second, it identifies clear and biologically meaningful spatial domains (top middle) that align well with the known cortical architecture. Third, the strong correspondence between the learned representations and manual layer annotations (top right) validates that the model captures genuine biological structure.

In contrast, STitch3D_batch_effect shows clear limitations without batch effect correction. The spots predominantly cluster by slice origin (bottom left), indicating that technical variation dominates the learned representations. The spatial domains (bottom middle) appear less well-defined, and the correspondence with manual annotations (bottom right) is notably weaker. These results emphasize that proper batch effect correction is crucial for integrating multiple tissue sections and extracting meaningful biological information from spatial transcriptomics data.

3.2.4 3D Spatial Domain Detection and Cell-type Distribution

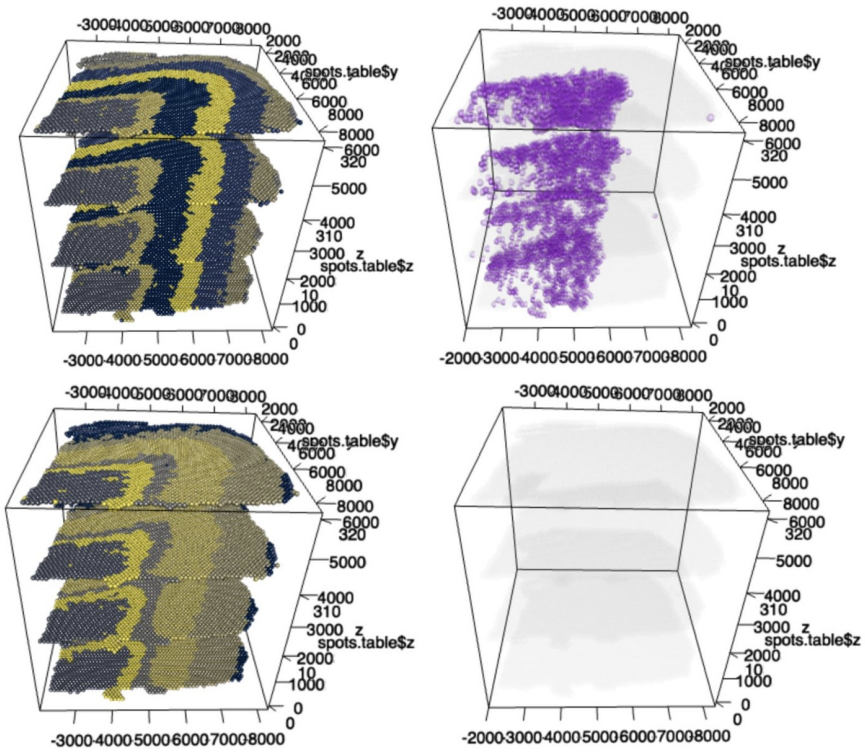

Figure 5: 3D visualization of STitch3D(Top Row) and STitch3D_no_batch_effect(Botton Row). Spatial domain detection results showing distinct clustering patterns(Left). Ex_8_L5_6 neuronal subtype distribution with proportion values higher then 20% highlighted in purple(Right).

The final visualization (Figure 5) provides comprehensive evidence for the superiority of the complete STitch3D model over the version without batch effect correction. In the upper panels, STitch3D demonstrates robust performance in both spatial domain detection and cell-type distribution mapping. The spatial domains are clearly delineated with distinct boundaries, indicating effective capture of tissue architecture. The Ex_8_L5_6 neuronal subtype distribution (purple regions) is well-defined and biologically meaningful.

In contrast, the lower panels reveal significant limitations of the model without batch effect correction. The spatial domain detection appears fragmented and inconsistent (bottom left), failing to capture coherent tissue organization. More strikingly, the cell-type distribution map (bottom right) shows almost complete absence of the Ex_8_L5_6 neuronal population (lack of purple regions), suggesting severe distortion of biological signals. This stark difference highlights the critical importance of batch effect correction in maintaining biological fidelity and achieving accurate 3D spatial reconstruction.

These results demonstrate that STitch3D's integrated batch effect correction is essential for reliable spatial transcriptomics analysis and accurate representation of cellular organization in three-dimensional space.

3.3 Experiments on Drosophila Embryo Whole Organism Atlas Dataset

We further evaluated STitch3D's performance on a larger real-world task: reconstructing a complete 3D Drosophila embryo model at 16-18h developmental stage, demonstrating its capability in whole organism spatial atlas reconstruction.

3.3.1 Data Preprocessing and Alignment

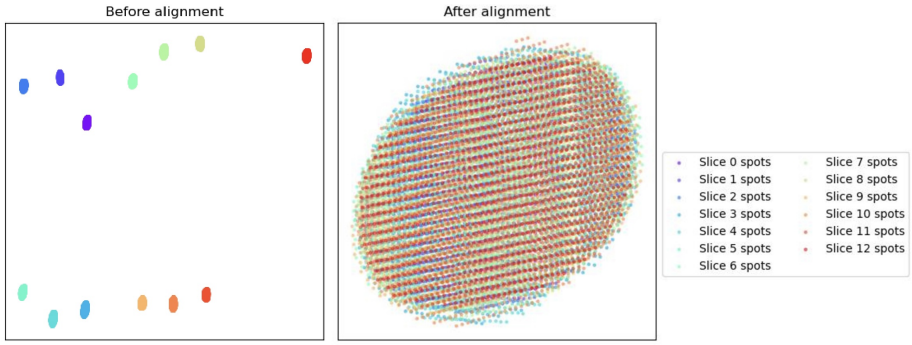

Figure 6: Slice alignment results for the Drosophila embryo dataset. (Left) Original slice positions showing significant misalignment between adjacent sections. (Right) Aligned slices after preprocessing using STitch3D's paste algorithm, demonstrating substantially improved spatial correspondence.

The alignment process (Figure 6) was particularly critical for this dataset due to the substantial variations between different slices of the Drosophila embryo. Without proper alignment, the different slices showed significant spatial inconsistencies that could compromise downstream analysis. STitch3D's paste algorithm successfully addressed this challenge by effectively aligning the slices while preserving the intricate anatomical features. The algorithm achieved remarkable improvement in spatial correspondence between adjacent sections, enabling accurate 3D reconstruction of the embryonic structure. This alignment step was fundamental for subsequent analyses, as it established a coherent spatial framework across all slices of the tissue.

The visualization demonstrates how the paste algorithm, with an alpha value of 0.2, effectively handled the complex 3D structure of the embryo, resulting in a well-integrated dataset where anatomical features align properly across different sections. This successful alignment was essential for accurate cell type mapping and spatial analysis in later steps.

3.3.2 Cell-type Deconvolution Comparison

The cell-type deconvolution analysis on the Drosophila embryo dataset (Figure 7) revealed interesting insights into the performance of both methods. While STitch3D_batch_effect showed improved performance on this large-scale dataset compared to previous experiments, possibly due to the increased data volume and stronger biological signals, it still fell short of STitch3D's capabilities in several aspects. Most notably, STitch3D successfully reconstructed the spatial distribution of CNS cell types, which STitch3D_no_batch_effect failed to achieve. STitch3D demonstrated superior performance in maintaining spatial coherence across adjacent slices and accurately capturing known anatomical structures. The method produced clear, biologically

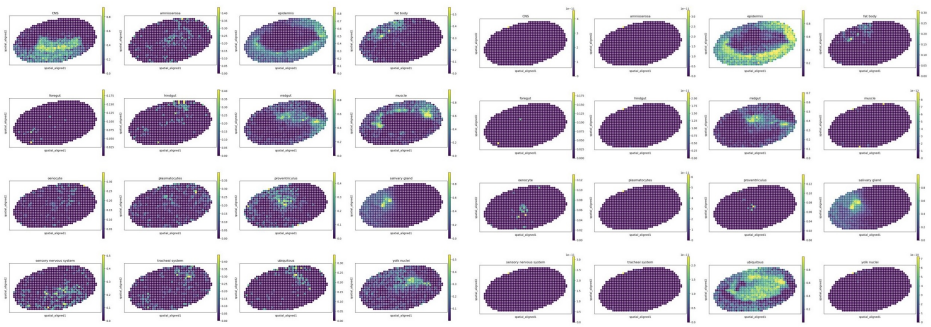

Figure 7: Comparison of cell-type deconvolution results between STitch3D (Left) and STitch3D_batch_effect (Right). Different colors represent distinct cell types, with intensity indicating proportion values.

meaningful delineations of tissue-specific cell populations, with consistent cell-type proportions that aligned well with established embryonic development patterns. In contrast, while STitch3D_no_batch_effect showed some improvement, it still exhibited less precise spatial patterns and struggled to maintain consistent cell-type assignments across different slices, particularly in regions requiring fine-grained cellular resolution.

3.3.3 3D Reconstruction of Key Anatomical Structures

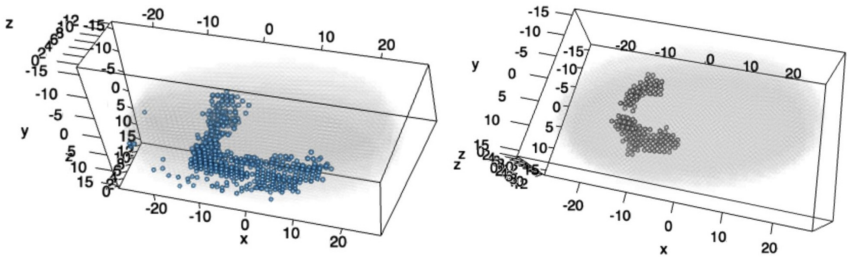

Figure 8: STitch3D's reconstructed 3D distributions of (Left) the central nervous system (CNS) and (Right) salivary gland. The reconstruction accurately captures the known morphology and positions of these structures.

STitch3D demonstrated excellent performance in reconstructing key anatomical structures in three dimensions (Figure 8). The reconstruction of the central nervous system (CNS) showed remarkable fidelity, with spatial organization that precisely matched known embryonic CNS morphology and positioning. Similarly, the salivary gland reconstruction accurately captured its distinctive structural features and anatomical location. These results highlight STitch3D's capability to faithfully reproduce complex biological structures while maintaining high accuracy in both morphological details and spatial relationships. The high-quality reconstructions provide strong validation of the method's effectiveness in preserving and representing critical anatomical features in three-dimensional space.

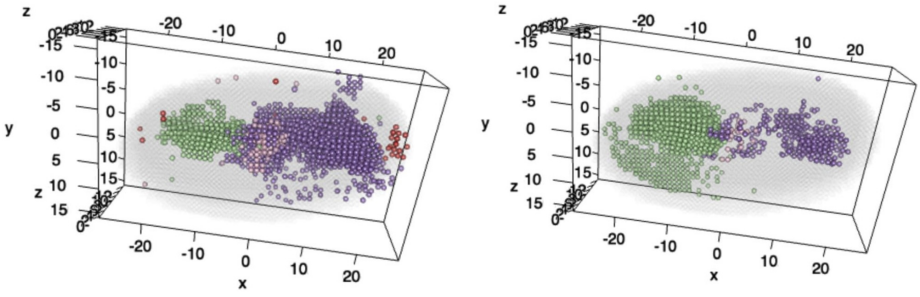

Figure 9: Visualization of estimated 3D distributions of the digestive system components using (Left) STitch3D and (Right) STitch3D_batch_effect. Spots with proportion values larger than 50% are shown. Different colors represent foregut, proventriculus, midgut and hindgut/anal pad regions.

3.3.4 Digestive System Reconstruction

The reconstruction of the digestive system components revealed remarkable differences between STitch3D and STitch3D_batch_effect methods, as shown in Figure 9. STitch3D demonstrated superior performance by successfully reconstructing the complete digestive tract with anatomically accurate spatial relationships. The method clearly delineated different regions from foregut to hindgut, including precise reconstruction of the proventriculus structure and the hindgut/anal pad region. In contrast, STitch3D_no_batch_effect produced notably inferior results, characterized by fragmented reconstruction of digestive tract components and poor spatial organization. Most significantly, it failed to properly reconstruct the hindgut/anal pad region and showed unclear boundaries between different digestive system components. These results highlight the importance of batch effect correction in achieving accurate 3D reconstruction of complex organ systems.

4 Advantages of STitch3D

4.1 Effective Information Representation

STitch3D's success is built upon a fundamental mathematical framework that effectively represents spatial transcriptomic data through the equation

$$X_{igs} = (P_i^s \cdot S_g) + \alpha_i^s + \gamma_g^s$$

, where X_{igs}represents the spatial transcriptomic expression matrix, P_i^s denotes the cell-type proportions matrix, S_g represents the reference cell-type expressions, and α_i^s and γ_g^s account for slice- and spot-specific effects and slice- and gene-specific effects, respectively. This decomposition effectively separates biological signals from technical variations while capturing spatial relationships through cell-type proportions and integrating reference knowledge via cell-type signatures.

A key innovation in STitch3D's architecture is its use of Graph Attention Networks (GAT) to project each spot's high-dimensional data into a 128-dimensional latent space representation. This projection is guided by the fundamental assumption that spots with similar biological characteristics should maintain their similarity in the latent space. The GAT architecture learns

these representations by adaptively aggregating information from neighboring spots, ensuring that the learned embeddings preserve both local and global spatial relationships. This latent representation serves as a crucial foundation for downstream tasks, particularly in cell-type proportion reconstruction and spatial domain clustering, where the reduced dimensionality and preserved biological relationships significantly enhance the model's ability to identify meaningful patterns and structures.

The effectiveness of this representation strategy is strongly validated by our experimental results, which demonstrate its necessity for accurate 3D reconstruction of spatial transcriptomics data. The combination of explicit batch effect modeling and learned latent representations enables STitch3D to capture complex biological relationships while effectively managing technical variations inherent in spatial transcriptomics data.

4.2 Uncertainty Handling

STitch3D addresses uncertainty through three key components: comprehensive batch effect modeling, dual loss functions, and a robust statistical framework.

The batch effect modeling explicitly incorporates two levels of technical variations:

- Slice- and spot-specific effects (α_i^s) capture local variations
- Slice- and gene-specific effects (γ_g^s) address systematic biases

The model employs two complementary loss functions for robust learning:

1. A Poisson-based Maximum Likelihood loss:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{g=1}^{G}[Y_{i,g}\log(l_i\lambda_{i,g}) - l_i\lambda_{i,g}]$$

   This function naturally handles the discrete nature of sequencing data and accounts for overdispersion in gene expression measurements.

2. An autoencoder regularization term:

$$R_{AE} = \frac{1}{N}\sum_{i=1}^{N}\|f_X(Z_i, s) - X_i\|_2$$

   Using a slice-specific decoder $f_X(\cdot, s)$, this term helps separate technical variations from biological signals while preserving meaningful patterns across slices.

The statistical framework combines these components with additional regularization terms to prevent overfitting. The Poisson distribution provides a theoretically sound approach for modeling count data, while the autoencoder component ensures biological interpretability. This integrated approach enables STitch3D to maintain robust performance across datasets of varying quality and complexity, effectively managing both technical variations and biological heterogeneity.

4.3 Computational Efficiency

STitch3D achieves remarkable computational efficiency through its innovative network architecture that combines Graph Neural Networks (GNNs) and Multi-Layer Perceptrons (MLPs). The use of GAT as the primary dimensionality reduction mechanism provides the model with exceptional flexibility and scalability, allowing it to handle datasets of varying sizes and complexities efficiently. This architecture enables effective parallel processing and GPU acceleration, significantly reducing computational time for large-scale analyses.

The model's efficiency stems from its thoughtful design choices: GNNs effectively capture spatial relationships between spots through adaptive attention mechanisms, while MLPs efficiently process high-dimensional gene expression data through optimized layer configurations. The graph attention mechanisms enable adaptive information aggregation, allowing the model to focus on the most relevant spatial relationships while ignoring noise. This combination results in a highly efficient computational framework that maintains accuracy while significantly reducing processing time.

Furthermore, the implementation leverages modern deep learning frameworks and GPU acceleration capabilities, enabling efficient processing of large-scale spatial transcriptomics datasets. The model's architecture is specifically optimized for parallel computation, allowing it to scale effectively with increasing dataset sizes while maintaining reasonable memory requirements. This computational efficiency, combined with the model's ability to handle batch effects and preserve biological signals, makes STitch3D particularly well-suited for analyzing complex, multi-slice spatial transcriptomics data at tissue or whole-organism scale.

5 Code Availability and Reproducibility

To ensure the reproducibility of our results, we have made all implementation code publicly available on GitHub[1]. The repository contains two main implementations:

- `STitch3D/`: The primary implementation of our proposed method with full functionality
- `STitch3D_no_batch_effect/`: An alternative implementation without batch effect correction

The code is primarily written in Python and presented in Jupyter Notebook format (85.3% of the codebase), making it easily accessible and executable. All experiments and results presented in this paper can be reproduced using these notebooks.

[1] `https://github.com/nosignalmxh/MATH-5472`